

AA 野生种花生 Ty1-copia 类反转座子 RT 基因的克隆 与序列分析¹

蔡铁城¹, 刘俊仙³, 张冲¹, 刘菁², 阳太亿², 蒋菁², 贺梁琼²,
韩柱强², 唐荣华², 庄伟建¹, 熊发前^{2*}

(1. 福建农林大学植物保护学院, 福州 350002; 2. 广西农业科学院
经济作物研究所, 南宁 530007; 3. 广西农业科学院甘蔗研究所/农业农村部广西甘蔗生物技
术与遗传改良重点实验室/广西甘蔗遗传改良重点实验室, 南宁 530007)

摘要: 该研究旨在克隆 Ty1-copia 类反转录转座子 RT 基因, 为分离花生属 Ty1-copia 类反转录转座子全长序列和研究其功能提供序列基础。根据 RT 基因的保守区设计简并引物, 以两份 AA 染色体组野生种花生“*Arachis duranensis*”为试材, 利用 PCR 扩增其基因组 DNA, 回收、克隆和测序目的条带后, 对所获得序列进行生物信息学分析。结果表明: (1) 目的条带大小约为 260 bp 大小, 分别从两份野生种花生材料中克隆到 41 条和 27 条 RT 基因序列, 68 条序列的长度变化范围为 256~270 bp, AT 所占比例范围为 55.86%~68.42%, AT 与 GC 比例范围为 1.27~2.17, 核苷酸序列间相似性范围为 49.8%~99.2%, 存在较高异质性。(2) 68 条序列被划分为 6 个家族, 家族 I 和 IV 为主要成分。(3) 68 条序列中的 19 条发生了无义突变, *Arachis duranensis* (PI219823) 要比 *Arachis duranensis* (PI262133) 的无义突变率高。(4) 氨基酸序列间相似性范围为 4.7%~100%, 呈现高度异质性。(5) 各家族中代表序列的蛋白质三级结构在整体构型上一致, 但在螺旋结构数、折叠结构数、转角数和氢键数上存在较大差别。(6) 序列间保守基序总体一致, 但也存在一定变异, 呈现一定异质性; 系统进化树将 68 条序列分为 10 类, 大部分序列都聚在 A 和 B 两大类中。(7) 另有部分 AA 染色体组野生种花生的 RT 基因序列与其他物种植物的 RT 基因序列亲缘关系较近, 推测不同物种植物之间可能发生过 Ty1-copia 类反转录转座子的横向传递。该研究为花生属基于 Ty1-copia 类反转录转座子的新分子标记开发和应用奠定基础。

关键词: 花生, Ty1-copia 类反转录转座子, 反转录酶, 野生种, 异质性

中图分类号: **文献标识码:**

Cloning and sequence analysis of reverse transcriptase of Ty1-copia-like retrotransposons in wild peanut species with

基金项目: 国家自然科学基金项目 (31960409, 32072103, 31960416); 广西自然科学基金项目 (2018GXNSFDA281027, 2018GXNSFDA294004, 2020GXNSFAA297081); 福建省科技厅农业引导性 (重点) 项目 (2018N0004); 广西农业科学院科技发展基金项目 (桂农科 2018YM06, 桂农科 2017JZ13, 桂农科 31960409, 桂农科 2021YT052) [Supported by the National Natural Science Foundation of China (31960409; 32072103; 31960416); Guangxi Natural Science Foundation Program (2018GXNSFDA281027; 2018GXNSFDA294004; 2020GXNSFAA297081); Science and Technology Foundation of Fujian Province of China (2018N0004); Guangxi Academy of Agricultural Sciences Fund Project (GNK2018YM06; GNK2017JZ13; GNK31960409; GNK2021YT052)].

作者简介: 蔡铁城(1983-), 硕士, 助理研究员, 主要从事花生遗传育种和分子生物学研究, (E-mail) caitiecheng1027@163.com。

***通信作者:** 熊发前, 博士生, 研究员, 主要从事花生种质资源和遗传育种及分子生物学研究, (E-mail) xfq2002@126.com。

AA genome

CAI Tiecheng¹, LIU Junxian³, ZHANG Chong¹, LIU Jing², YANG Taiyi², JIANG Jing², HE Liangqiong², HAN Zhuqiang², TANG Ronghua², ZHUANG Weijian¹, XIONG Faqian^{2*}

(1. College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2. Cash Crops Research Institute, Guangxi Academy of Agricultural Sciences,

Nanning 530007, China; 3. Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences/Sugarcane Research Center, Chinese Academy of Agricultural Sciences/Key Laboratory of Sugarcane Biotechnology and Genetic Improvement (Guangxi), Ministry of Agriculture and Rural Affairs/Guangxi Key Laboratory of Sugarcane Genetic Improvement, Nanning 530007)

Abstract: The purpose of this study was to clone the RT genes of Ty1-copia-like retrotransposons, and to provide sequences basis for isolating the full-length sequences of Ty1-copia-like retrotransposons and studying their function in genus *Arachis*. Degenerate primers were designed according to the conserved region of RT genes, the genomic DNA of wild peanut species "*Arachis duranensis*" with AA genome was amplified by PCR using the degenerated primers. The amplified targeted bands were recovered, cloned and sequenced, and then sequences were analyzed through bioinformatics strategy. The results were as follows: (1) The amplified targeted bands were all about 260 bp in size. Forty-one and twenty-seven RT genes sequences were cloned from the two wild peanut species respectively. The length of sixty-eight sequences varied from 256 bp to 270 bp. The proportion of AT, AT/GC and the similarity between nucleotide sequences ranged from 55.86% to 68.42%, 1.27% to 2.17%, and 49.8% to 99.2% respectively, showing a higher heterogeneity. (2) The sixty-eight sequences were divided into six families, family I and family IV were the main components. (3) Nineteen of sixty-eight sequences had nonsense mutations, and *Arachis duranensis* (PI219823) had a higher nonsense mutations rate than *Arachis duranensis* (PI262133). (4) The similarity between amino acid sequences ranged from 4.7% to 100%, showing high heterogeneity. (5) The tertiary structures of proteins representing sequences in each family were basically similar in overall configuration, but there were great differences in the number of helix structures, folding structures, turns and hydrogen bonds. (6) The conserved motifs among sequences were generally consistent, but there were also some variations, showing a certain degree of heterogeneity. The phylogenetic tree divided sixty-eight sequences into ten classes. Most of sequences are clustered in A and B classes. (7) Some of RT genes sequences from two wild peanut species with AA genome were closely related to RT genes sequences from other plant species, which indicated that there might be transposon horizontal transmission between them. This study laid a foundation for the development and application of new molecular markers based on Ty1-copia-like retrotransposons in genus *Arachis*.

Key words: peanut, Ty1-copia-like retrotransposons, reverse transcriptase, wild species, heterogeneity

花生被誉为“长生果”，全世界有 106 个国家种植。花生是我国主要食用油原料，我国是世界上重要的花生生产国。当前我国选育出的花生品种主要是利用传统杂交育种，但存在周期长、效率低、目的性不强等缺点，而分子育种可以加速育种进程，但是缺少简单实用高效的 DNA 分子标记。在花生上，传统分子标记技术想检测出丰富多态性十分困难（熊发前等，2010；王强等，2010；Xiong et al., 2011），虽然近几年在花生上也有利用 SNP 标记进行关联及连锁分析的研究报道（Zhang et al., 2017；Han et al., 2018；Wang et al., 2018），但当前在花生上使用最广泛的是 SSR 标记，但能在任意两个栽培种花生品种间检测出 DNA 多态性的 SSR 标记引物对还比较匮乏（熊发前等，2010；Xiong et al., 2011）。

LTR 反转录转座子主要包括 Ty1-*copia* 和 Ty3-*gypsy* 两大类 (Kumar & Bennetzen, 1999; Feschotte et al., 2002; Bonchev & Parisod, 2013), 这两类 LTR 反转录转座子中的 RT 基因序列都可以通过简并 PCR 技术扩增克隆 (Voytas et al., 1992; Kumekawa et al., 1999)。LTR 反转录转座子具有的普遍性、高拷贝、高度异质性和插入位点多态性等特性使其非常适合开发分子标记。

但到目前为止, 在花生上尚未见到基于 LTR 反转录转座子开发分子标记的研究报道。而分离和鉴定 LTR 反转录转座子是分子标记开发利用的前提。花生 LTR 反转录转座子的研究报道稀少, Nielen 等先后分离出花生 Ty3-*gypsy* 类反转录转座子的 *FIDEL* 和 Ty1-*copia* 类反转录转座子的 *Matita*, 对 *FIDEL* 和 *Matita* 的特性和作用进行了分析 (Nielen et al., 2010, 2012)。笔者曾系统对花生 LTR 反转录转座子和 MITE 转座子的分离及其应用的国内外研究现状及进展进行了归纳 (熊发前等, 2017)。

本研究拟从两份 AA 染色体组野生种花生材料“*Arachis duranensis*”中克隆 Ty1-*copia* 类反转录转座子 RT 基因, 分析其序列特征和多样性, 为分离花生属 Ty1-*copia* 类反转录转座子全长序列和研究其功能提供序列基础, 为花生属基于 Ty1-*copia* 类反转录转座子的新分子标记开发和应用奠定基础。

1 材料与方法

1.1 供试材料

在广西农业科学院武鸣里建科研基地的花生野生圃里随机摘取两份 AA 染色体组野生种花生材料“*Arachis duranensis*” (PI262133) 和“*Arachis duranensis*” (PI219823) 各 5 株健康植株的顶端芯叶进行混合。

1.2 基因组 DNA 的提取

花生高质量基因组 DNA 的提取采用改良 CTAB 法 (熊发前等, 2019)。

1.3 RT 基因的 PCR 扩增

上游引物为: RTp1: 5'-ACNGCNTTYTNCAYGG-3', 下游引物为: RTp2: 5'-ARCATRTCRTCACRTA-3', 其中, R=A/G, Y=C/T, N=A/T/C/G (Kumar et al., 1997)。PCR 扩增体系和扩增程序以及 PCR 产物的分离检测参考报道文献进行 (熊发前等, 2019)。

1.4 PCR 产物的回收、克隆及测序

参考报道文献进行 (阳太亿等, 2019)。

1.5 RT 基因的序列分析

序列相似性检索、序列统计分析、序列图及 Logo 图的生成、蛋白二级结构与三级结构预测、蛋白三级结构的转角数和氢键数统计、保守基序预测等参考报道文献进行 (阳太亿等, 2019)。运用 MEGA6.0 软件的邻接法 (No. of differences 模型) 构建系统进化树, 自展值设置为 1000, 所用其他物种植物的 Ty1-*copia* 类反转录转座子 RT 基因序列信息见表 1。

表 1 部分其他物种植物的 Ty1-*copia* 类反转录转座子 RT 基因序列信息

Table 1 The amino acid sequences information of reverse transcriptase of Ty1-*copia*-like retrotransposons from some other plant species

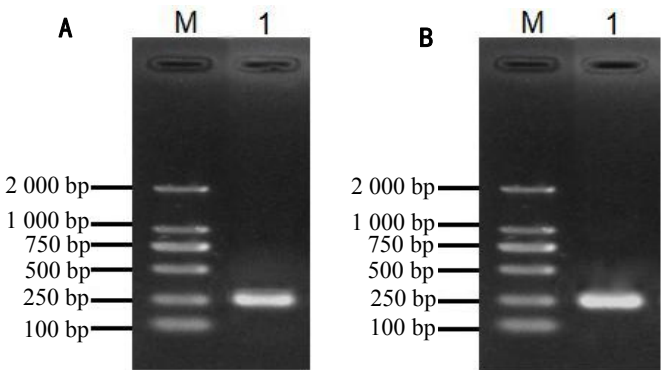
| 科 | 物种 | 登录号 |
|---------------|-----------------------------------|---------------|
| Family | Species | Accession No. |
| 茄科 Solanaceae | 烟草 <i>Nicotania tabacum</i> | AAA03507.1 |
| | 马铃薯 <i>Solanum tuberosum</i> | CAA13067.1 |
| | 辣椒 <i>Capsicum annuum</i> | AFS89521.1 |
| | 番茄 <i>Lycopersicon esculentum</i> | AAC34611.1 |
| 豆科 Fabaceae | 鹰嘴豆 <i>Cicer arietinum</i> | CAD59770.1 |

| | | |
|--------------------|---------------------------------|------------|
| | 绿豆 <i>Vigna radiata</i> | AAT90460.1 |
| | 绿豆 <i>Vigna radiata</i> | AAT90479.1 |
| | 绿豆 <i>Vigna radiata</i> | AAT90494.1 |
| | 大豆 <i>Glycine max</i> | E47759 |
| | 水稻 <i>Oryza sativa</i> | AAA33902.1 |
| 禾本科 Poaceae | 玉米 <i>Zea mays</i> | AAK84849.1 |
| | 大狗尾草 <i>Setaria faberi</i> | AAL36472.1 |
| | 苹果 <i>Malus domestica</i> | ABS11062.1 |
| 蔷薇科 Rosaceae | 梅花 <i>Prunus mume</i> | ABF57071.1 |
| | 李子 <i>Prunus salicina</i> | AGX45518.1 |
| 藜科 Chenopodiaceae | 藜麦 <i>Chenopodium quinoa</i> | AEX61031.1 |
| 十字花科 Brassicaceae | 拟南芥 <i>Arabidopsis thaliana</i> | S71291 |
| | 油菜 <i>Brassica napus</i> | AAA32987.1 |
| 山茶科 Theaceae | 野茶树 <i>Camellia sinensis</i> | CAJ09751.1 |
| 松科 Pinaceae | 欧洲云杉 <i>Picea abies</i> | CAA11921 |
| 石蒜科 Amaryllidaceae | 石蒜 <i>Lycoris radiata</i> | CAA11484.1 |
| 五加科 Araliaceae | 西洋参 <i>Panax quinquefolius</i> | ABU94811.1 |
| 葡萄科 Vitaceae | 葡萄 <i>Vitis vinifera</i> | CAN67451.1 |

2 结果与分析

2.1 RT 基因的 PCR 扩增及测序

对 AA 染色体组野生种花生材料“*Arachis duranensis*” (PI262133) 和“*Arachis duranensis*” (PI219823) 的基因组 DNA 进行 PCR 扩增, 结果显示, 在 2 份花生材料中都扩增出了大小约 260 bp 的特异条带 (图 1)。将目的条带进行回收、克隆和测序, 从 *Arachis duranensis* (PI262133) 和 *Arachis duranensis* (PI219823) 中分别获得了 42 条和 38 条序列。利用 DNAMAN 软件去除相同序列, 利用 NCBI 数据库对序列进行同源性分析进而去除非目标序列, 最后从 *Arachis duranensis* (PI262133) 和 *Arachis duranensis* (PI219823) 中分别获得了 41 条和 27 条目标序列, 并分别命名为 AdRT1-X 和 AdRT2-X (表 1)。对这些 RT 基因序列进行多重比对 (图 2), 利用 weblogo 生成了序列 logo 图以展示每个位置上碱基的保守性 (图 3)。



(A) M: DL2000 marker, 1: *Arachis duranensis*(PI262133); (B) M: DL2000 marker, 1: *Arachis duranensis*(PI219823)。

图 1 *Arachis duranensis* 中 Ty1-copia 类反转录转座子 RT 基因的 PCR 扩增
Fig. 1 PCR amplification of reverse transcriptase of Ty1-copia-like retrotransposons from *Arachis duranensis*

from *Arachis duranensis*

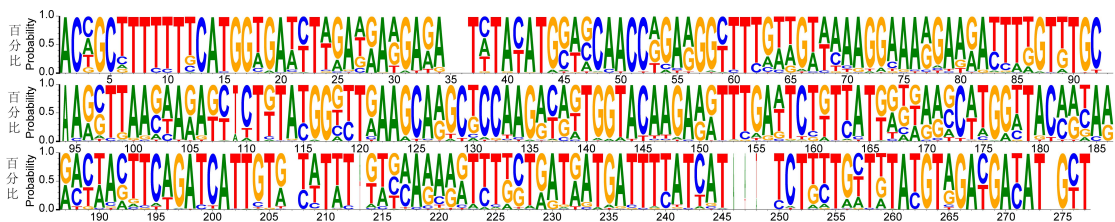


图3 *Arachis duranensis* 中 Ty1-*copia* 类反转录转座子 RT 基因序列的比对 logo

Fig. 3 Alignment logo of reverse transcriptase of Ty1-*copia*-like retrotransposons amplified from *Arachis duranensis*

2.2 RT 基因序列分析

利用 BioEdit 软件对 RT 基因序列进行统计分析, 结果显示, 所有序列长度都在 256~270 bp 之间, 存在缺失或插入突变。在 *Arachis duranensis* (PI262133) 的 41 条序列中, AdRT1-47 的序列长度最短, 为 256 bp, AdRT1-11 的序列长度最长, 为 267 bp, 长度为 266 bp 的序列占总序列数的 90.24% (表 1); A、T、C、G 数量变化范围分别为 65~88、78~94、29~61 和 49~67, AT 所占比例范围为 55.86%~68.42%, AT 与 GC 比例为 1.27~2.17 (表 2); 核苷酸序列间相似性范围为 50%~99.2%, 其中, AdRT1-26 与 AdRT1-25 之间的相似性最高, 达 99.2%, AdRT1-47 与 AdRT1-7 以及 AdRT1-47 与 AdRT1-29 之间的相似性最低, 均为 50%, 氨基酸序列间相似性范围为 4.7%~100%。在 *Arachis duranensis* (PI219823) 的 27 条序列中, AdRT2-24 的序列长度最短, 为 257 bp, AdRT2-7 的序列长度最长, 为 270 bp, 长度为 266 bp 的序列占总序列数的 74.07% (表 1); A、T、C、G 数量变化范围分别为 62~93、71~93、32~56 和 66~68, AT 所占比例范围为 57.20%~67.29%, AT 与 GC 比例为 1.34~2.06 (表 2); 核苷酸序列间相似性范围为 49.8%~99.2%, 其中, AdRT2-3 与 AdRT2-4 之间的相似性最高, 达 99.2%, AdRT2-13 与 AdRT2-27 之间的相似性最低, 为 49.8%, 氨基酸序列间相似性范围为 14.6%~100%。

表 2 “*Arachis duranensis*”中 Ty1-*copia* 类反转录转座子 RT 基因序列的基本信息

Table 2 Basic information of reverse transcriptase of Ty1-*copia*-like retrotransposons amplified from *Arachis duranensis*

| 序列编号 Sequence No. | 大小 Size (bp) | AT 含量 AT content (%) | AT/GC | 登录号 Accession No. |
|----------------------|-----------------|-------------------------|-------|----------------------|
| AdRT1-1 | 266 | 63.16 | 1.71 | MK830202 |
| AdRT1-2 | 266 | 63.53 | 1.74 | MK830203 |
| AdRT1-4 | 266 | 62.78 | 1.69 | MK830204 |
| AdRT1-6 | 266 | 63.53 | 1.74 | MK830205 |
| AdRT1-7 | 266 | 67.29 | 2.06 | MK830206 |
| AdRT1-8 | 266 | 62.78 | 1.69 | MK830207 |
| AdRT1-9 | 266 | 63.16 | 1.71 | MK830208 |
| AdRT1-11 | 267 | 64.42 | 1.81 | MK830209 |
| AdRT1-12 | 266 | 62.41 | 1.66 | MK830210 |
| AdRT1-13 | 266 | 62.03 | 1.63 | MK830211 |
| AdRT1-14 | 266 | 63.53 | 1.74 | MK830212 |
| AdRT1-15 | 266 | 62.78 | 1.69 | MK830213 |
| AdRT1-16 | 262 | 59.92 | 1.50 | MK830214 |
| AdRT1-17 | 266 | 62.41 | 1.66 | MK830215 |
| AdRT1-19 | 266 | 61.65 | 1.61 | MK830216 |
| AdRT1-21 | 266 | 62.78 | 1.69 | MK830217 |

| | | | | |
|-----------------|---------|-------------|-----------|----------|
| AdRT1-22 | 266 | 62.78 | 1.69 | MK830218 |
| AdRT1-23 | 265 | 63.02 | 1.70 | MK830219 |
| AdRT1-24 | 266 | 63.16 | 1.71 | MK830220 |
| AdRT1-25 | 266 | 62.78 | 1.69 | MK830221 |
| AdRT1-26 | 266 | 62.41 | 1.66 | MK830222 |
| AdRT1-28 | 266 | 62.78 | 1.69 | MK830223 |
| AdRT1-29 | 266 | 68.42 | 2.17 | MK830224 |
| AdRT1-30 | 266 | 62.03 | 1.63 | MK830225 |
| AdRT1-31 | 266 | 62.78 | 1.69 | MK830226 |
| AdRT1-32 | 266 | 63.53 | 1.74 | MK830227 |
| AdRT1-33 | 266 | 63.91 | 1.77 | MK830228 |
| AdRT1-34 | 266 | 64.66 | 1.83 | MK830229 |
| AdRT1-35 | 266 | 62.41 | 1.66 | MK830230 |
| AdRT1-36 | 266 | 61.28 | 1.58 | MK830231 |
| AdRT1-37 | 266 | 62.41 | 1.66 | MK830232 |
| AdRT1-38 | 266 | 63.91 | 1.77 | MK830233 |
| AdRT1-39 | 266 | 62.78 | 1.69 | MK830234 |
| AdRT1-40 | 266 | 63.53 | 1.74 | MK830235 |
| AdRT1-41 | 266 | 62.41 | 1.66 | MK830236 |
| AdRT1-42 | 266 | 65.41 | 1.89 | MK830237 |
| AdRT1-43 | 266 | 62.78 | 1.69 | MK830238 |
| AdRT1-45 | 266 | 63.16 | 1.71 | MK830239 |
| AdRT1-46 | 266 | 64.66 | 1.83 | MK830240 |
| AdRT1-47 | 256 | 55.86 | 1.27 | MK830241 |
| AdRT1-48 | 266 | 64.66 | 1.83 | MK830242 |
| 变化范围 | | | | |
| Variation range | 256~267 | 55.86~68.42 | 1.27~2.17 | - |
| AdRT2-1 | 266 | 62.78 | 1.69 | MK830471 |
| AdRT2-2 | 266 | 63.53 | 1.74 | MK830472 |
| AdRT2-3 | 266 | 63.16 | 1.71 | MK830473 |
| AdRT2-4 | 266 | 63.16 | 1.71 | MK830474 |
| AdRT2-5 | 267 | 66.29 | 1.97 | MK830475 |
| AdRT2-7 | 270 | 64.07 | 1.78 | MK830476 |
| AdRT2-9 | 266 | 62.78 | 1.69 | MK830477 |
| AdRT2-10 | 262 | 60.69 | 1.54 | MK830478 |
| AdRT2-11 | 266 | 62.78 | 1.69 | MK830479 |
| AdRT2-13 | 266 | 63.91 | 1.77 | MK830480 |
| AdRT2-14 | 266 | 62.78 | 1.69 | MK830481 |
| AdRT2-15 | 266 | 63.91 | 1.77 | MK830482 |
| AdRT2-16 | 266 | 62.78 | 1.69 | MK830483 |
| AdRT2-18 | 266 | 62.41 | 1.66 | MK830484 |
| AdRT2-19 | 263 | 62.36 | 1.66 | MK830485 |
| AdRT2-21 | 267 | 66.29 | 1.97 | MK830486 |
| AdRT2-24 | 257 | 57.20 | 1.34 | MK830487 |
| AdRT2-27 | 266 | 67.29 | 2.06 | MK830488 |

| | | | | |
|-----------------|---------|-------------|-----------|----------|
| AdRT2-37 | 266 | 63.53 | 1.74 | MK830489 |
| AdRT2-38 | 266 | 66.54 | 1.99 | MK830490 |
| AdRT2-39 | 266 | 61.65 | 1.61 | MK830491 |
| AdRT2-40 | 266 | 63.16 | 1.71 | MK830492 |
| AdRT2-41 | 266 | 62.78 | 1.69 | MK830493 |
| AdRT2-42 | 266 | 63.16 | 1.71 | MK830494 |
| AdRT2-43 | 266 | 64.29 | 1.80 | MK830495 |
| AdRT2-45 | 267 | 62.92 | 1.70 | MK830496 |
| AdRT2-48 | 266 | 61.65 | 1.61 | MK830497 |
| 变化范围 | 257~270 | 57.20~67.29 | 1.34~2.06 | - |
| Variation range | | | | |

2.3 RT 基因核苷酸序列聚类分析

利用 MEGA6.0 软件对 RT 基因序列进行聚类分析（图 4）。遗传进化树显示，68 条序列被划分为 6 个家族，家族 I 包含 28 条序列，其中，18 条来自 *Arachis duranensis*(PI262133)，10 条来自 *Arachis duranensis*（PI219823），所含序列数占总序列数的 41.18%；家族 II 包含 9 条序列；家族 III 只包含 AdRT1-29，该序列与家族 I 亲缘关系较远而单独聚为一类，造成该序列单独为一类的原因可能是碱基替换；家族 IV 包含 23 条序列，其中有 17 条来自 *Arachis duranensis*（PI262133），6 条来自 *Arachis duranensis*（PI219823），所含序列数占总序列数的 33.82%；家族 V 包含 3 条序列，家族 VI 包含 4 条序列，家族 V 和家族 VI 与另外 4 个家族遗传距离较大、亲缘关系较远，这两个家族中的序列存在碱基缺失的现象。

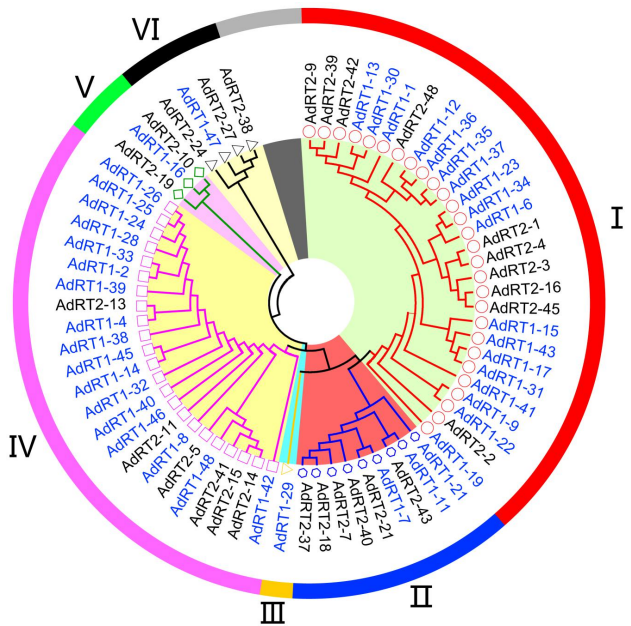
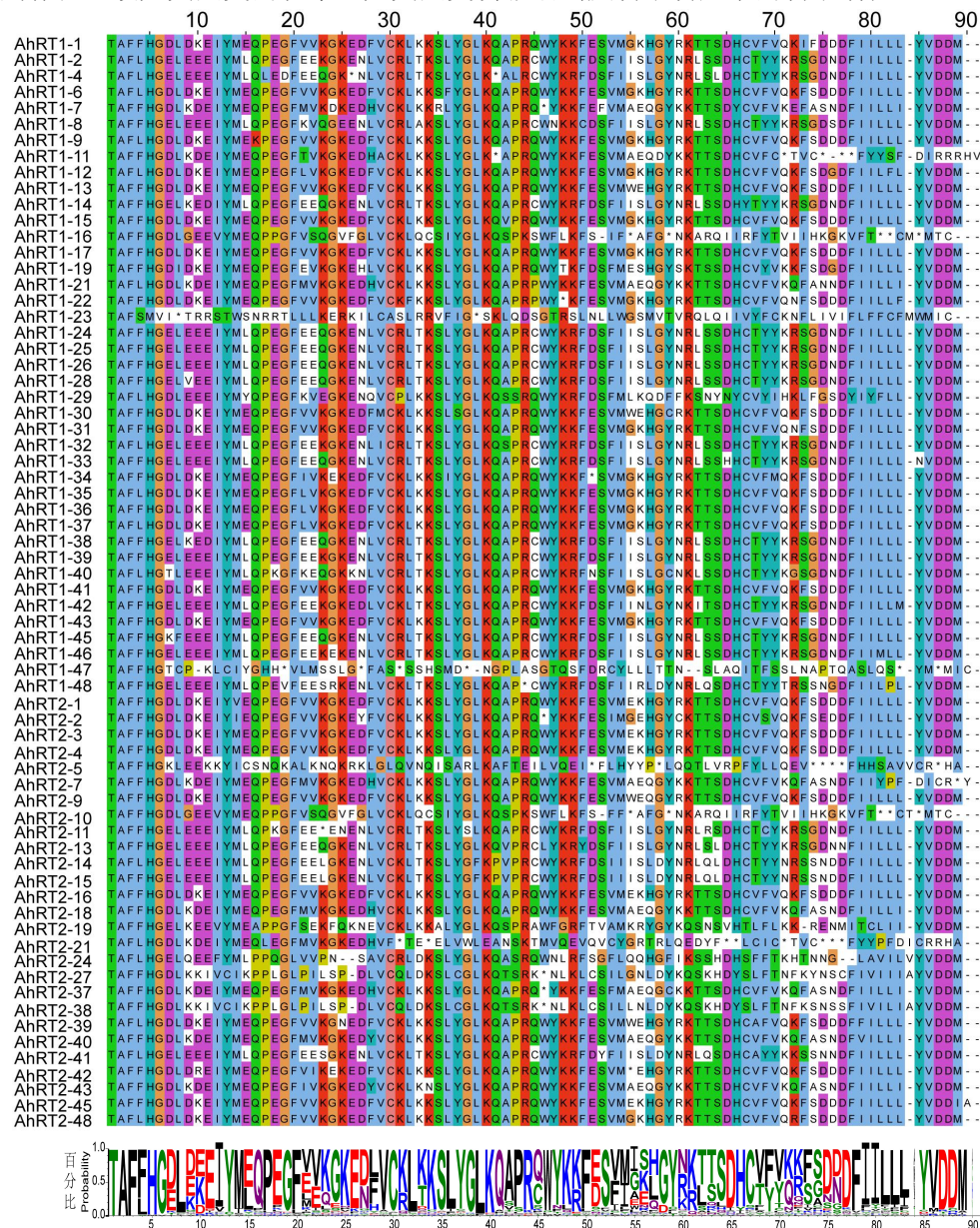


图 4 *Arachis duranensis* 中 Ty1-copia 类反转录转座子 RT 基因的遗传进化树
Fig. 4 Phylogenetic tree of reverse transcriptase of Ty1-copia-like retrotransposons amplified from *Arachis duranensis*

2.4 RT 基因氨基酸序列分析

利用 MEGA6.0 软件对 RT 基因氨基酸序列进行分析（图 5）。结果显示，68 条序列中有 19 条发生无义突变，其中，9 条序列来自 *Arachis duranensis*（PI262133），占该材料序列总数的 21.95%，10 条序列来自 *Arachis duranensis*（PI219823），占该材料序列总数的 37.04%，就无义突变发生率来说，*Arachis duranensis*（PI219823）要比 *Arachis duranensis*（PI262133）高。无义突变在 2 份花生材料中的具体表现为：AdRT2-21 发生了 8 个无义突变，分别在第 31、34、65、66、71、75、76、77 个氨基酸处；AdRT2-5 发生了 7 个无义突

变, 分别在第 51、58、74、75、76、77、87 个氨基酸处; AdRT1-47 发生了 6 个无义突变, 分别在第 18、26、30、37、79、82 个氨基酸处; AdRT1-11 发生了 5 个无义突变, 分别在第 41、71、75、76、77 个氨基酸处; AdRT1-16 发生了 5 个无义突变, 分别在第 54、58、80、81、84 个氨基酸处; AdRT2-10 发生了 5 个无义突变, 分别在第 54、58、81、82、84 个氨基酸处; AdRT1-4 发生了 2 个无义突变 (第 26、41 个氨基酸处); AdRT1-23 发生了 2 个无义突变 (第 8、39 个氨基酸处); AdRT1-7、AdRT1-22、AdRT1-34、AdRT1-48、AdRT2-7、AdRT2-11、AdRT2-27、AdRT2-37、AdRT2-38 和 AdRT2-42 都只发生了 1 个无义突变; 部分序列存在连续无义突变的现象, 无义突变会影响到反转录转座子的转录活性。



*表示终止密码子; - 表示优化联配产生的缺口。

* indicates stop codes; - indicates gaps introduced for optimal alignment.

图 5 *Arachis duranensis* 中 Ty1-copia 类反转录转座子 RT 基因氨基酸序列的多重比对及其 logo

Fig. 5 Multiple alignment and its logo of amino acid sequences of reverse transcriptase of Ty1-copia-like retrotransposons amplified from *Arachis duranensis*

2.5 RT 基因的蛋白结构预测

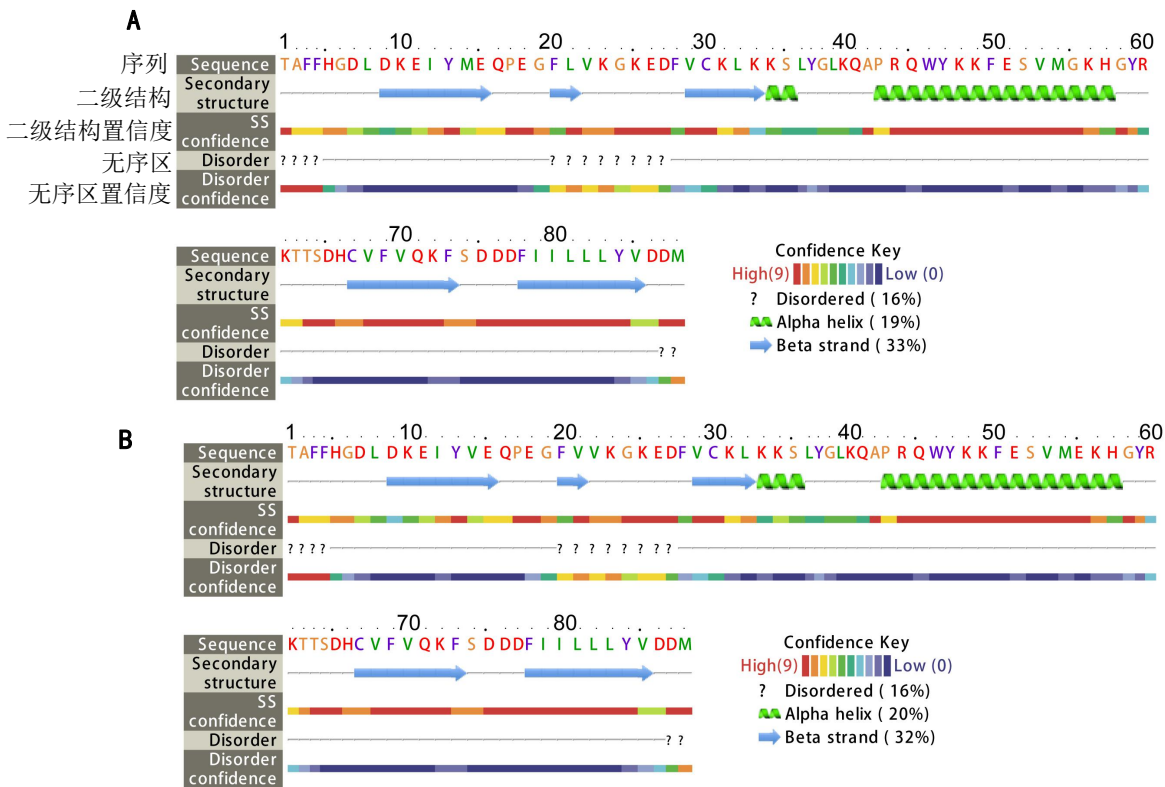
将获得的 68 条 Ty1-*copia* 类反转录转座子 RT 基因序列统一翻译成氨基酸后，根据核苷酸聚类结果，分别选择两份花生材料中每个家族中的代表序列，利用在线程序 Phyre2 预测其蛋白质的二级结构和三级结构（表 3，图 6，图 7），代表序列蛋白三级结构匹配覆盖度最高的模板为 d1hara、c4rs7R、d1sqwa1、c5xvnM，置信度均为 16.5~86.2，其中只有 d1hara 属于逆转录酶家族，其余蛋白均为匹配到逆转录蛋白模板。二级结构包含 2~3 个 α -螺旋和 5 个 β -折叠；三级结构包含 2~6 个转角和 9~30 个氢键，还存在 1 个明显的螺旋结构和 2 个不明显的折叠结构。

表 3 *Arachis duranensis* 中 Ty1-*copia* 类反转录转座子 RT 基因的蛋白结构信息
Table 3 Protein structure information of reverse transcriptase of Ty1-*copia*-like retrotransposons amplified from *Arachis duranensis*

| 家族 Family | 序列号 Sequence No. | α -螺旋数 α -helix number ^a | β -折叠数 β -strand number ^a | 转角数 Corner number ^b | 氢键数 Hydrogen bond number ^b |
|--------------|------------------------|---|--|--------------------------------------|---|
| I | AdRT1-36 | 2 | 5 | 5 | 25 |
| | AdRT2-1 | 2 | 5 | 6 | 28 |
| II | AdRT1-7 | 2 | 5 | 5 | 30 |
| | AdRT2-18 | 3 | 5 | 5 | 30 |
| III | AdRT1-29 | 2 | 5 | 5 | 29 |
| IV | AdRT1-46 | 2 | 5 | 6 | 29 |
| | AdRT2-13 | 2 | 5 | 5 | 30 |
| V | AdRT1-16 | 2 | 5 | 2 | 16 |
| | AdRT2-19 | 3 | 5 | 6 | 29 |
| VI | AdRT1-47 | 2 | 5 | 4 | 9 |
| | AdRT2-27 | 2 | 5 | 3 | 23 |

注：^a的数据结果是根据图 6 的蛋白质二级结构统计得出的；^b的数据结果是根据图 7 的蛋白质三级结构统计得出的。

Note: The results of “a” are based on the protein secondary structure of Fig. 6; while those of “b” are based on the protein tertiary structure of Fig. 7.



A: AdRT1-36; B: AdRT2-1。下同。

A: AdRT1-36; B: AdRT2-1. The same below.

图 6 *Arachis duranensis* 中 Ty1-copia 类反转录转座子 RT 基因的蛋白二级结构

Fig. 6 Protein secondary structure of reverse transcriptase of Ty1-copia-like retrotransposons amplified from *Arachis duranensis*

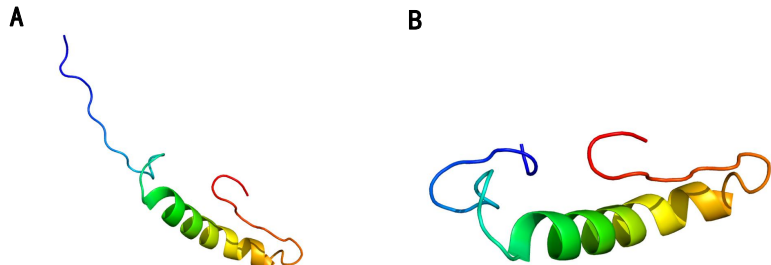


图 7 *Arachis duranensis* 中 Ty1-copia 类反转录转座子 RT 基因的蛋白三级结构

Fig. 7 Protein tertiary structure of reverse transcriptase of Ty1-copia-like retrotransposons amplified from *Arachis duranensis*

2.5 RT 基因保守基序预测

RT 基因保守基序预测结果显示, 68 条序列共存在 11 种保守基序, 其中, 有 57 条序列同时包含 motif 1、motif 2 和 motif 3, 占总序列数的 83.82%, 说明这 3 种保守基序是 AA 染色体组野生种花生 Ty1-copia 类反转录转座子 RT 基因的主要保守基序, 这也从保守基序角度表明两个 AA 染色体组野生种花生 Ty1-copia 类反转录转座子 RT 基因具有非常高的保守性与相似性; AdRT2-27 和 AdRT2-38 同时包含 motif 3、motif 4 和 motif 6, motif 4 和 motif 6 这两个保守基序在位置上与 motif 1 和 motif 2 相同, 但氨基酸的排列和组成并不相同; AdRT1-11 包含 motif 1、motif 3 和 motif 7; AdRT1-16 和 AdRT2-10 同时包含 motif 1 和 motif 5, 在系统进化树中, 这 2 条序列为同一类; AdRT2-19 包含 motif 1 和 motif 10, AdRT2-24 包含 motif 8、motif 9 和 motif 10, AdRT2-21 包含 motif 1 和 motif 7, 这 3 条序列在系统进化树中均单独归类; AdRT1-47 包含 motif 8 和 motif 9, AdRT2-5 包含 motif 7、motif 8 和 motif

11, AdRT1-23 只包含 motif 11, 这 3 条序列在系统进化树中归为同一类, 其中, motif 8 和 motif 11 的长度较短, 位于序列上游部分。部分保守基序在所克隆序列中出现的频率较低且长度较短, 说明这些序列在进化过程中发生了突变(图 8)。

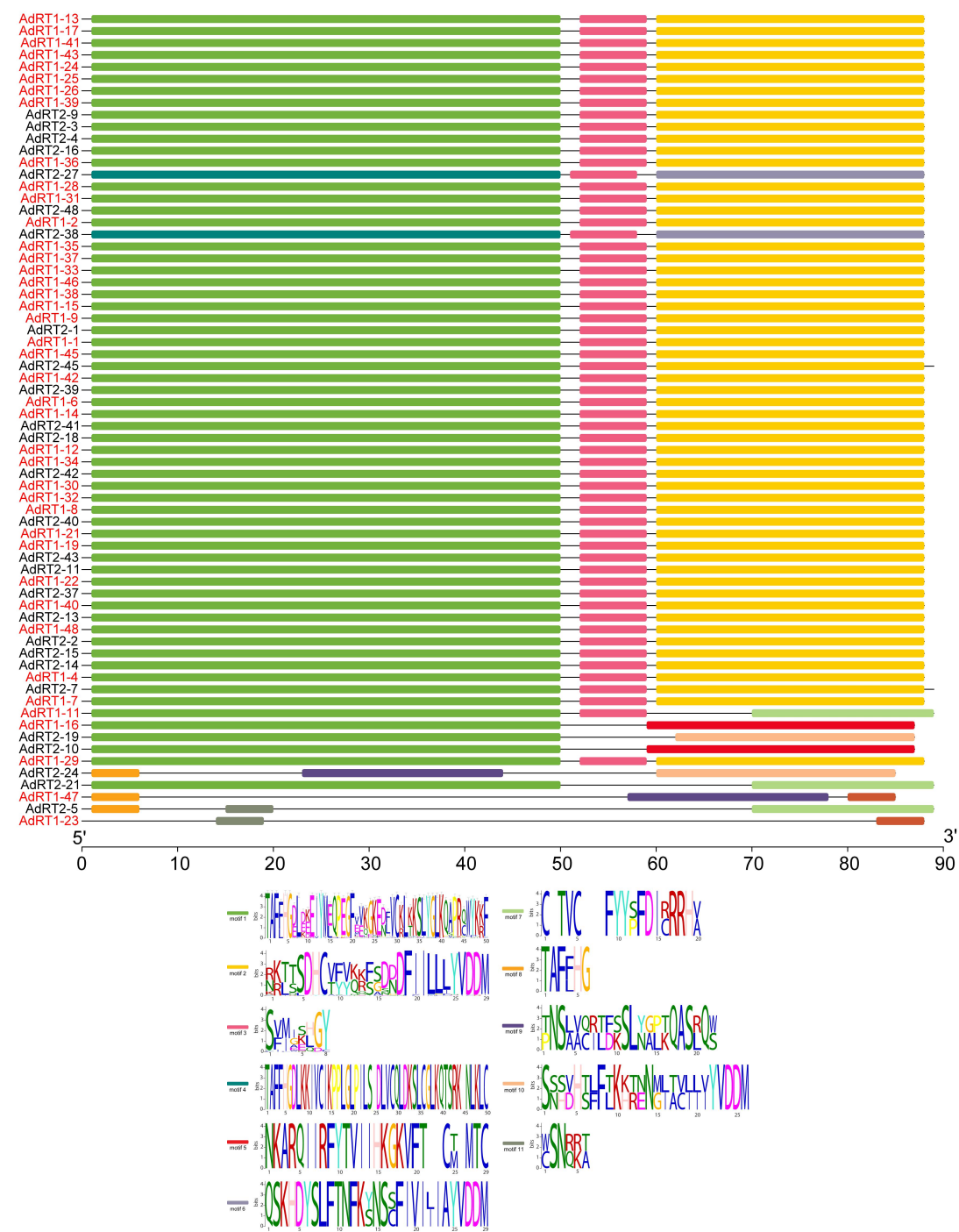


图 8 *Arachis duranensis* 中 Ty1-*copia* 类反转录转座子 RT 基因保守基序预测

Fig. 8 Conservative motifs prediction of reverse transcriptase of Ty1-*copia*-like retrotransposons amplified from *Arachis duranensis*

2.6 RT 基因序列系统进化树构建

构建系统进化树(图 9), 所有 RT 基因序列可分为 10 类, 大部分 RT 基因序列都聚在 A 和 B 两大类中, 表明本研究中 AA 染色体组野生种花生的 RT 基因序列具有相当高的保守性与相似性。A 类包含 20 条来自 *Arachis duranensis*(PI262133)和 15 条来自 *Arachis duranensis*

(PI219823) 的序列, 这 35 条序列与来自葡萄 (*Vitis vinifera*, CAN67451.1)、绿豆 (*Vigna radiata*, AAT90460.1、AAT90479.1)、鹰嘴豆 (*Cicer arietinum*, CAD59770.1)、马铃薯 (*Solanum tuberosum*, CAA13067.1)、西洋参 (*Panax quinquefolius*, ABU94811.1)、野茶树 (*Camellia sinensis*, CAJ09751.1)、李子 (*Prunus salicina*, AGX45518.1)、苹果 (*Malus domestica*, ABS11062.1)、烟草 (*Nicotiana tabacum*, AAA03507.1) 的序列之间具有较高相似性, 亲缘关系较近。B 类包含 17 条来自 *Arachis duranensis*(PI262133)和 5 条来自 *Arachis duranensis* (PI219823) 的序列, 说明这 22 条序列之间具有较高相似性。C 类只有 AdRT1-29, 该序列与来自油菜 (*Brassica napus*, AAA32987.1)、梅花 (*Prunus mume*, ABF57071.1)、大豆 (*Glycine max*, E47759)、藜麦 (*Chenopodium quinoa*, AEX61031.1)、番茄 (*Lycopersicon esculentum*, AAC34611.1)、大狗尾草 (*Setaria faberi*, AAL36472.1)、水稻 (*Oryza sativa*, AAA33902.1) 和玉米 (*Zea mays*, AAK84849.1) 的序列之间具有较高相似性, 亲缘关系较近。D 类中是来自绿豆、欧洲云杉和石蒜的 3 条序列, 这 3 条序列与大多数 AA 染色体组野生种花生和其它物种植物的序列亲缘关系较远。E 类只包含 2 条序列, 分别是来自 AA 染色体组野生种花生的 AdRT2-24 和拟南芥的 S71291, 说明这 2 条序列之间亲缘关系最近。F 类~J 类中的 3 条和 7 条序列分别来自 *Arachis duranensis* (PI262133) 和 *Arachis duranensis* (PI219823), 均不包含其它物种植物的序列。

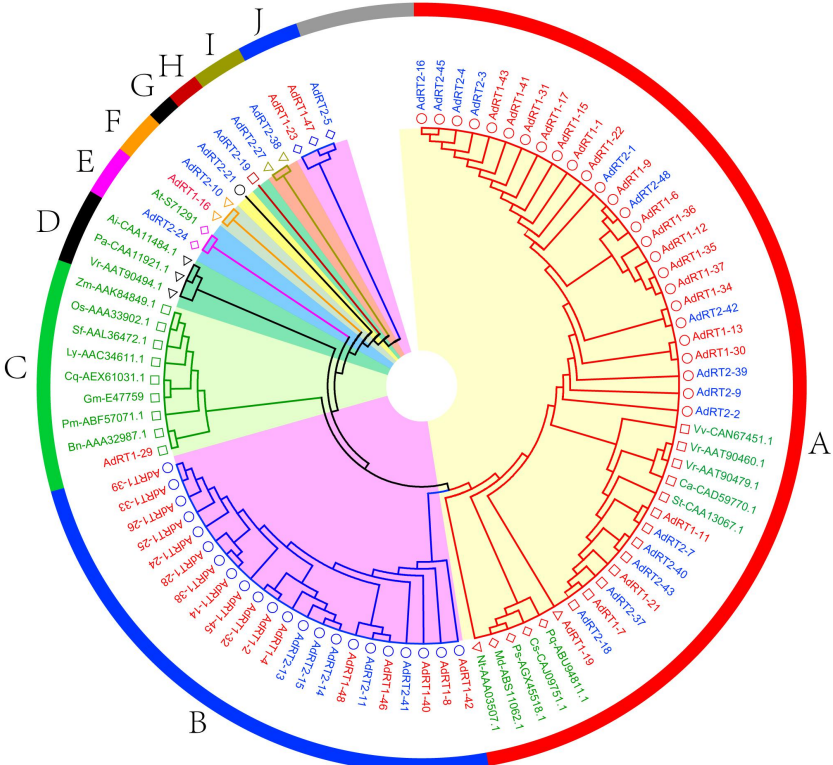


图 9 *Arachis duranensis* 与部分其他物种植物 Ty1-copia 类反转录转座子 RT 基因序列的系统进化树

Fig. 9 Phylogenetic tree of reverse transcriptase of Ty1-copia-like retrotransposons from *Arachis duranensis* and some other plant species

3 讨论与结论

本研究用同一简并引物从同一种野生花生种质中克隆获得的 Ty1-copia 类反转录转座子 RT 基因的核苷酸和氨基酸序列在序列长度、序列组成及序列相似性等方面存在较大差异和

多态性,表明同一种野生花生种质内同一类群反转录转座子存在较高异质性。另外,两份野生种花生材料中 RT 基因序列均富含 AT 碱基,富含 AT 导致呈现较高异质性;RT 基因核苷酸序列间相似性呈现较高异质性;有 19 条氨基酸序列发生了无义突变,导致呈现较高异质性;RT 基因氨基酸序列间相似性呈现高度异质性;68 条 RT 基因序列间保守基序也呈现一定异质性;家族 V 和家族 VI 中的代表序列在螺旋结构、折叠结构、转角数、氢键数上与其他家族代表序列存在的较大差别,呈现较高异质性和多态性。总之,两份野生种花生材料中 RT 基因序列在 AT 碱基含量、核苷酸序列间相似性、氨基酸序列间相似性、氨基酸无义突变率、保守基序、蛋白质二级结构及三级结构上均呈现异质性。

家族 I~IV 中代表序列的蛋白质三级结构在整体构型上基本类似,但在螺旋结构数、折叠结构数、转角数和氢键数上存在较大差别,家族 V 中的 AdRT1-16 只存在 2 个明显的螺旋结构,其转角数和氢键数也明显少于其他序列,家族 VI 中的 AdRT1-47 的氢键数远远少于其它序列,其转角数也少于其他序列,且其螺旋结构较短,AdRT2-27 相较于其他序列存在 2 个明显的折叠结构和 1 个不明显的折叠结构,其转角数和氢键数都少于其他序列,推测这些差别可能会影响到 Ty1-*copia* 类反转录转座子的拷贝数、转录活性及转座效率等。

遗传进化树显示,6 个家族中的 I 和 IV 是主要家族,表明 AA 染色体组野生种花生 Ty1-*copia* 类反转录转座子 RT 基因序列具有非常高的保守性与相似性。另外,各家族内部成员越复杂,序列相似性越高,存在有转录活性的反转录转座子的可能性也越大,其转座发生的时间可能越近(Tang et al., 2005),由此推测,家族 I 和家族 IV 很有可能是存在具有转录活性的 Ty1-*copia* 类反转录转座子的家族,存在的历史也更为久远。

系统进化树显示,A 类中的 35 条 AA 染色体组野生种花生 RT 基因序列与葡萄、绿豆、鹰嘴豆、马铃薯、西洋参、野茶树、李子、苹果、烟草的序列之间相似性较高;C 类中的 AdRT1-29 与来自油菜、梅花、大豆、藜麦、番茄、大狗尾草、水稻、玉米的序列之间相似性较高;E 类中的 AdRT2-24 与拟南芥的 S71291 之间的相似性较高,推测 AA 染色体组野生种花生 Ty1-*copia* 类反转录转座子曾可能与这些物种植物间发生过横向传递。F 类~J 类中的 RT 基因序列均来自 AA 染色体组野生种花生,且与 AA 染色体组野生种花生和其他物种植物的 RT 基因序列遗传距离最大,亲缘关系最远,反映这几类中的 RT 基因序列在起源和进化上可能较为古老,特异性比较强,有可能为 *Arachis duranensis* (PI262133) 和 *Arachis duranensis* (PI219823) 所特有。

总之,本研究从 AA 染色体组野生种花生中成功克隆到 Ty1-*copia* 类反转录转座子 RT 基因序列,对花生属基于 LTR 反转录转座子的分子标记开发及花生分子育种具有重要意义,将为下一步分离其全长序列、研究其转录和转座活性及功能提供序列基础。

参考文献:

- BONCHEV G, PARISOD C, 2013. Transposable elements and microevolutionary changes in natural populations[J]. Mol Ecol Resour, 13(5): 765-775.
- FESCHOTTE C, JIANG N, WESSLER SR, 2002. Plant transposable elements: where genetics meets genomics[J]. Nat Rev Genet, 3(5): 329-341.
- HAN SY, YUAN M, CLEVENGER JP, et al., 2018. A SNP-based linkage map revealed QTLs for resistance to early and late leaf spot diseases in peanut (*Arachis hypogaea* L.)[J]. Front Plant Sci, 9: 1012.
- KUMAR A, BENNETZEN JL, 1999. Plant retrotransposons[J]. Annu Rev Genet, 33(1): 479-532.
- KUMEKAWA N, OHTSUBO E, OHTSUBO H, 1999. Identification and phylogenetic analysis of gypsy-type retrotransposons in the plant kingdom[J]. Genes Genet Syst, 74(6): 299-307.
- NIELEN S, CAMPOS-FONSECA F, LEAL-BERTIOLI S, et al., 2010. FIDEL-a retrovirus-like retrotransposon and its distinct evolutionary histories in the A- and B-genome components of

- cultivated peanut[J]. Chrom Res, 18(2): 227-246.
- NIELEN S, VIDIGAL BS, LEAL-BERTIOLI SCM, et al., 2012. *Matita*, a new retroelement from peanut: characterization and evolutionary context in the light of the *Arachis* A-B genome divergence[J]. Mol Genet Genom, 287(1): 21-38.
- TANG YM, MA YZ, LI LC, et al., 2005. Identification and characterization of reverse transcriptase domain of transcriptionally active retrotransposons in wheat genome[J]. J Integr Plant Biol, 47(5): 604-612.
- VOYTAS DF, CUMMINGS MP, KONIECZNY A, et al., 1992. *Copia*-like retrotransposons are ubiquitous among plants[J]. Proc Natl Acad Sci USA, 89(15): 7124-7128.
- WANG LF, ZHOU XJ, REN XP, et al., 2018. A major and stable QTL for bacterial wilt resistance on chromosome B02 identified using a high-density SNP-based genetic linkage map in cultivated peanut Yuanza 9102 derived population[J]. Front Genet, 9: 652.
- WANG Q, ZHANG XY, TANG FS, et al., 2010. Construction of genetic linkage map of peanut (*Arachis hypogaea* L.) based on SRAP markers[J]. Chinese journal of oil crop science, 32(3): 374-378. [王强, 张新友, 汤丰收, 等, 2010. 基于 SRAP 分子标记的栽培种花生遗传连锁图谱构建[J]. 中国油料作物学报, 32(3): 374-378.]
- XIONG FQ, JIANG J, ZHONG RC, et al., 2010. Application of SCoT molecular markers in genus *arachis*[J]. Acta Agronomica Sinica, 36(12): 2055-2061. [熊发前, 蒋菁, 钟瑞春, 等, 2010. 目标起始密码子多态性 (SCoT) 分子标记技术在花生属中的应用[J]. 作物学报, 36(12): 2055-2061.]
- XIONG FQ, LIU JX, HE LQ, et al., 2017. Recent advances on the development and utilization of molecular markers based on LTR retrotransposons and MITE transposons from peanut (*Arachis hypogaea* L.) [J]. Mol Plant Breed, 15(2): 640-647. [熊发前, 刘俊仙, 贺梁琼, 等, 2017. 花生 LTR 和 MITE 转座子及其分子标记开发利用研究进展[J]. 分子植物育种, 15(2): 640-647.]
- XIONG FQ, LIU JX, LIU J, et al., 2019. Comparative analysis and application of five improved CTAB extraction methods for peanut DNA[J]. Mol Plant Breed, 17(7): 2207-2216. [熊发前, 刘俊仙, 刘菁, 等, 2019. 花生 DNA 的五种改良 CTAB 提取方法的比较分析及其应用[J]. 分子植物育种, 17(7): 2207-2216.]
- XIONG FQ, ZHONG RC, HAN ZQ, et al., 2011. Start codon targeted polymorphism for evaluation of functional genetic variation and relationships in cultivated peanut (*Arachis hypogaea* L.) genotypes[J]. Mol Biol Rep, 38(5): 3487-3494.
- YANG TY, LIU JX, LIU J, et al., 2019. Cloning and analysis of reverse transcriptase of Ty1-*copia*-like retrotransposons in *Arachis monticola*[J]. Shandong Agric Sci, 51(9): 9-20, 27. [阳太亿, 刘俊仙, 刘菁, 等, 2019. 四倍体野生种花生 Ty1-*copia* 类逆转座子逆转录酶基因的克隆与分析[J]. 山东农业科学, 51(9): 9-20, 27.]
- ZHANG XG, ZHANG JH, He XY, et al., 2017. Genome-wide association study of major agronomic traits related to domestication in peanut[J]. Front Plant Sci, 8: 1611.